Oil Palm Detection via Deep Transfer Learning

Isis Bonet
EIA University
Envigado, Colombia
isis.bonet@eia.edu.co

Fabio Caraffini
De Montfort University
Leicester, UK
fabio.caraffini@dmu.ac.uk

Alejandro Peña
EIA University
Envigado, Colombia
juan.pena@eia.edu.co

Alejandro Puerta
EIA University
Envigado, Colombia
alejandro.puerta@eia.edu.co

Mario Gongora
De Montfort University
Leicester, UK
mgongora@dmu.ac.uk

Abstract—This article presents an intelligent system using deep learning algorithms and the transfer learning approach to detect oil palm units in multispectral photographs taken with unmanned aerial vehicles. Two main contributions come from this piece of research. First, a dataset for oil palm units detection is carefully produced and made available online. Although being tailored to the palm detection problem, the latter has general validity and can be used for any classification application. Second, we designed and evaluated a state-of-the-art detection system, which uses a convolutional neural network to extract meaningful features, and a classifier trained with the images from the proposed dataset. Results show outstanding effectiveness with an accuracy peak of 99.5% and a precision of 99.8%. Using different images for validation taken from different altitudes the model reached an accuracy of 97.5% and a precision of 98.3%. Hence, the proposed approach is highly applicable in the field of precision agriculture.

Index Terms—Classification, Convolutional neural networks, Deep learning , Oil palm, Multispectral image processing

I. INTRODUCTION

Worldwide, the oil palm industry leads the supply of oils and fats. It is a very dynamic sector due to the breadth of uses of palm oil in different products, from cooking to cleaning products, as well as in special greases and lubricants, personal hygiene and cosmetics, production of biodiesel and electrical energy [1]. In this light, oil palms are valuable and play an important role in the economy of some countries. As an example, they are considered by the Colombian Government as one of the main agricultural products to promote the substitution of illegal crops and achieve job creation in the countryside, thus allowing for a sustainable peace process evolution.

To support the management of the crop, low-cost unmanned aerial vehicles (UAVs) [2], also known as drones, are being used for various purposes such as pesticide spraying. An emerging trend involves their use for acquiring high-resolution aerial images, to cut the costs as compared to using satellite images. This approach is convenient as it easily allows for the generation of both multispectral and thermal images, with a very accurate resolution [2], [3] and controlled acquisition process and equipment on the UAVs. With small drones with spectral cameras, it is possible to take photos and do tasks such as aerial mapping, plant health monitoring, and weed detection, among others. All these tasks involve image processing tasks, where the first step is the segmentation of the image and the detection of the crop units (individual plants).

Application domains are varied, but mostly related to precision farming and agriculture, e.g. from prevention of fires [4], to the classification of different plant species [5], [6], early diagnosis of plant diseases [7], [8], or even employed for counting out single plant units and in the census of live animals [9].

Most of these processes first require isolating a single palm tree from those in its surrounding and the elimination of vegetation with similar characteristics, as e.g. grass or bushes, from the image. In a nutshell, individual palm tree identification is the first step to perform, and the performance of the subsequent processes depend on the quality and accuracy of this identification phase. Hence, improving upon the palm detection process is key to optimally perform several other tasks.

A. Advances in Crop Units Detection

The rise of Artificial Intelligence (AI) models for image analysis such as deep learning and associated methods has attracted the attention of researchers in the field of precision agriculture [10] with a particular emphasis on embedding it in the oil palm sector where an informed use of AI in the oil production process can lead to major improvements with high economical, environmental and suitability impact [8], [11], [12]. Indeed, several studies are analysing spectral images with AI tools for the identification of crop units [13], [14].

Currently, a great deal of research is being carried out for the detection of oil palm units using these latest deep learning models, including transfer learning methods [15], [16] and specific neural systems such as Convolutional Neural Netowk (CNN) [17], [18]. Many of them use high-resolution satellite images to train CNN models capable of detecting palm units and return their exact number in a a given area [19]–[21]. However, sometimes satellite images are too expensive or not available in real-time to analyse these crops. Hence, alternatives making use of UAVs to collect multi-spectral photographs have been investigated and has been shown to provide suitable images to detect palm trees with CNNs [22] and a Support Vector Machines (SVMs) [23].

B. Contributions and Structure of this work

In line with the most best performing research reported in the literature, this work proposes a novel intelligent system for

the automatic spatial identification of oil palm trees based on:

- the collection of multispectral photographs with an UAV;
- the generation of a data set suitable for designing and training AI solutions specifically tailored for the identification of oil palm units;
- the use of advanced AI methods, i.e. CCNs and SVMs, to extract the features and identify the units.

Hence, the contribution of this research is two-fold. First, an original data set was carefully prepared as explained in section II-B, and made available online at [24] for the research community to easily work on images that are otherwise difficult to collect (due to the location and accessibility of oil palms plantations). Second, a palm detection system based on "deep transfer learning" is designed and evaluated.

To present our work, the remainder of this paper is structured as follows:

- section II describes the data collection process, its pre-processing for generating a training dataset, the AI "tools" involved in this work, the methods employed in this study and evaluation methods to assess the validity of the proposed system;
- section III present the intelligent system for oil palm unit detection;
- section IV presents and discusses the results obtained;
- section V concludes this work with final considerations and identifies possible ways forward for future improvements and new research lines.

II. MATERIALS AND METHODS

A. Collecting multi spectral photographs

Photos were taken using a "Dji Phantom 3" drone equipped with a "Parrot sequoia" multispectral camera, as shown in Fig. 1, at 4 different spectrum bands: Green $(550 - 590nm)$, Red $(660 - 700nm)$, Red Edge $(735 - 745nm)$ and Near Infrared$(790 - 830nm)$ with a resolution of 1280×960 for each band (except for RGB images, where the resolution is 4608×3456). Multiple flights, each one covering 6400 palm units, were performed at different times of the day and at different altitudes, i.e. 10m, 20m, 30m, 40m and 50m, to take photographs under various scenarios. It must be noted that some of the most common problems in image processing for precision agriculture are: 1) the same plant can look different in two separate photos due to different sunlight conditions during the day, 2) it can be difficult to understand the age of a plant from the relative size seen in photos taken at different altitudes; hence, it is useful to acquire images under multiple configurations of these factors to address these issues in any study. In total 400 photos were taken at each frequency band, which means that $4 \times 400 = 1600$ images were collected.

All the images were taken in the Santa Barbara plantation owned by Unipalma S.A. [25], a company located in Meta province of Colombia (4° 13' 33" N, 73° 14' 50"W) which supported the collection of the images for this study.

(a) Dji Phantom 3   (b) Parrot sequoia

Fig. 1. Collecting photographs in Santa Barbara (Unipalma [25]) with our UAV (a) equipped with a multispectral camera (b). More images are available in [26].

B. Building the dataset

To build the training data base we carefully selected 8 of the Near InfraRed (NIR) images taken at an altitude of 50m. These are available at [27] – one is shown as an example in Fig. II-B. Subsequently, these images were cropped into multiple sub-images having different sizes: those containing only a complete single oil palm unit, an example is shown in Fig. II-B, were grouped to form the "poitive" classification case (i.e. a palm is detected); conversely, those containing more the one units or incomplete ones, an example can be seen in Fig. II-B, were grouped to form the "negtive" classification case (i.e. a whole palm tree is not detected). The process of accurately selecting positive cases was performed manually and double-checked to guarantee the appropriateness of the images. Similarly, negative cases were produced by cropping rectangular portions of the original images at random positions and with different aspect ratios. In both cases, images were inspected and labelled before being associated with the positive or negative class.

The two image galleries were compressed in two separate archives, one containing 356 positive instances while the other one a total of 938 negative instances, both stored in the online dataset repository [24]. The latter, can be used to train various classification models.

C. Sliding window

The "sliding window" technique [28], [29] is one of the most used algorithms to segment the image in object detection. In its simplest implementation, it consists of a rectangle of fixed width and height slid through an image from left to right and from top to bottom with a fixed step size.

Adaptive sliding windows do also exist, see e.g. [30], but are not required for this study since random window sizes are more suitable. To insure this, a square sliding window with initial size of 117×117 pixels is run N= 3 times with a fixed step equal to 32 pixels to process new input images as shown in Fig.3 and generate sub-images to be subsequently classified. Each time the sliding window algorithm is run, the window size randomly varies in the range $117 - 181$ pixels (for images taken at an altitude of 50m). It must be pointed
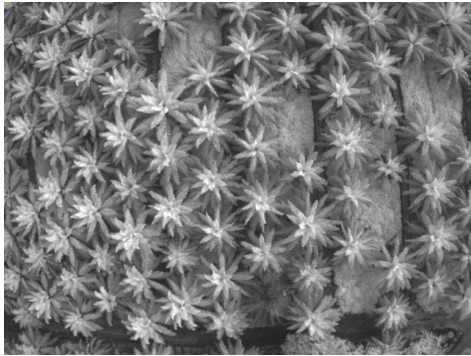
(a) NIR original image
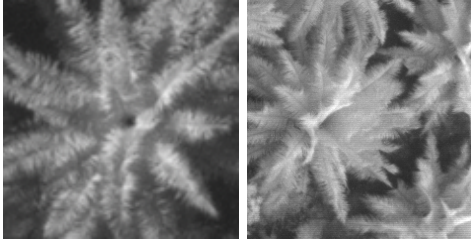
(b) Positive   (c) Negative

Fig. 2. Generating the training dataset: a NIR images (a) is decomposed to generate positive (b) and negative (c) instances. The complete gallery of the original NIR images and the whole dataset (positive plus negative instances) are available on the online repository [26].

out that if the proposed system is validated with images taken at different altitudes, these ranges should be adapted accordingly to achieve optimal classification performances. For lower altitudes we suggest increasing the range while we suggest to decrease it for higher altitude values. As an example, we empirically found the range $309 - 661$ to be optimal for an altitude of 10m. This addresses the issue of relative size of a palm unit depending on the altitude of the acquisition flight, thus arising the need of having altitude-customised boundaries for the sliding window size.

D. Convolutional Neural Network

The CNN [17] is a deep learning model consisting of a feed-forward artificial neural network that has convolution and pooling layers, for feature extraction, and fully connected layers, for classification tasks [31], [32]. CNNs are suitable for many applications in a variety of fields and are particularly suited to the field of computer vision, for which state-of-the-art pre-trained deep neural network architectures have successfully been designed [18]. Some of the most prominent CNN structures are known under the names VGG-16, VGG-19, Inception V3, ResNet-50 and Xception.

This work makes use of a VGG-16 network without its last layer (i.e. the fully connected layer equipped with a "softmax" activation function); its topology is described in Table I. Further technical information are available from the documentation of the popular "keras" package [33].

TABLE I
VGG-16 MODEL

Layer type	Output Shape	Parameters number
input (InputLayer)	(224,224,3)	0
block1_conv1 (Conv2D)	224,224,64)	1792
block1_conv2 (Conv2D)	(224,224,64)	36928
block1_pool (MaxPooling2D)	(112,112,64)	0
block2_conv1 (Conv2D)	(112,112,128)	73856
block2_conv2 (Conv2D)	(112,112,128)	147584
block2_pool(MaxPooling2D)	(56,56,128)	0
block3_conv1 (Conv2D)	(56,56,256)	295168
block3_conv2 (Conv2D)	(56,56,256)	590080
block3_conv3 (Conv2D)	(56,56,256)	590080
block3_pool(MaxPooling2D)	(28,28,256)	0
block4_conv1 (Conv2D)	(28,28,512)	1180160
block4_conv2 (Conv2D)	(28,28,512)	2359808
block4_conv3 (Conv2D)	(28,28,512)	2359808
block4_pool(MaxPooling2D)	(14,14,512)	0
block5_conv1 (Conv2D)	(14,14,512)	2359808
block5_conv2 (Conv2D)	(14,14,512)	2359808
block5_conv3 (Conv2D)	(14,14,512)	2359808
block5_pool (MaxPooling2D)	(7,7,512)	0
flatten (Flatten)	(25088)	0
fc1 (Dense)	(4096)	102764544
fc2 (Dense)	(4096)	16781312
predictions (Dense)	(1000)	4097000
Total parameters:		14714688
Trainable parameters:		14714688
Non-trainable parameters:		0

This CNN plays a key role in the proposed system as it responsible for extracting features from the images to be classified, as graphically shown in Fig. 3.

E. Transfer learning for feature extraction

Transfer Learning (TL) is a machine learning methodology to exploit knowledge acquired during a previous a problem, and used to address a new and different problem [34], [35]. Hence, it can be used for training new models based on previous training process, in particular when 1) training data availability is poor or 2) time constraints are preventing from performing an full training process. Therefore models pre-trained over extensive datasets with similar features to the new problem can be used as a high performance starting point.

The TL approach is particularly efficient in image processing, given that it has been observed that deep neural networks with different sets of natural images seems to learn similar features [35], [36]. This means that many image classification problems, including the palm tree detection task, can be dealt by starting with pre-trained methods. Therefore, the previously discussed pre-trained CNN is used off-the-shelf in this study to extract features from input images, as shown in figure 3.

Actual		Predicted	
		Positive	**Negative**
	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

F. Support Vector Machine

The SVM is a supervised machine learning model used for classification and regression [37]. The original SVM algorithm was a linear classifier for a two-class problem, separating the data with two parallel hyperplanes such that the distance between them is maximised. The region between the two hyperplanes is in jargon referred to as "margin", while the samples falling in the margin are referred to as the "support vectors" which are separated in two classes by the "maximum margin hyperplane" [38].

Although the SVM makes a linear separation in the space to perform the classification, kernel functions can be used to map data into a higher dimensional space, thus allowing also for non-linear classifications. The kernel functions are inner products in the feature space and the most widely used are linear, polynomial and Gaussian radial basis [39].

In this work, a SVM is used as a classifier to detect positive cases.

G. Evaluation methods

The "k-fold" cross-validation method is used to validate the proposed model. Cross-validation helps us to find the unknown fit parameters in the classification methods, as well as to estimate the prediction error of the final model. So we can estimate how the model will behave in new cases [40].

Furthermore, to evaluate its performances the most significant metrics based on the confusion matrix (see Table II) are used as well. These are defined as described below in (1), (2), (3) and (4):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

III. THE PROPOSED DEEP TRANSFER LEARNING MODEL

The methods described in section II are employed together as outlined in Fig. 3 to form the proposed oil palm unit detection system.

This process consists of four main steps:

1) **Preparation of a suitable dataset**: source images are not suitable inputs for a detection system and require processing to be transformed into smaller "fragments" that can be labelled, in this case according to the presence/absence of a single whole oil palm unit, before being fed into a classifier. In this study this is achieved with the methods described in section II-B which led to the creation of the dataset for palm trees detection archived in [24]. This dataset is suitable for binary classification in which a "positive" outcome is the detection of an oil palm tree while a "negative" one consists in detection of an incomplete palm, a different object or plant, or multiple palms.

2) **Features extraction**: to achieve optimal classification, in this step all images from the prepared dataset are transformed into a more informative format highlighting their key characteristics. This process, aka feature extraction, is efficiently performed with TL using a VGG-16 CNN without its last layer (as explained in section II-E) and pre-trained over the large-scale ImageNet database [41]. Advantages of using ImageNet include the availability of various training objects and shapes that lead to a fine tuning of the VGG-16 parameters (see Table I) that could not be obtained otherwise. Furthermore, rotated version of training objects are included to make sure that performances of the CNN are not dependent on rotation of the inputs images. In fact this database has been thought to perform object recognition, amongst a total of 1000 available objects, and this is why pre-trained CNN architectures terminate with a "softmax" layer that is responsible classifying the objects in their classes. However, this part is not of interest for this study, given that we want to use the CNN for feature extraction purposes only. We later deal with our particular classification to have higher detection accuracy. Therefore, we exploit the VGG-16 architecture until its penultimate layer, which returns features. It is worth mentioning that some authors suggest to eliminate the so-called "last block", consisting of three layers, to perform a more general feature extraction process for tasks that do not require object detection [42]. However, identifying palm trees can be seen as a very specific case of object detection and we empirically observed better performances, for this specific application domain, by removing only the last layer.

Under the configuration described previously, temporary copies of the labelled images from the previous step are then resized into the same 224×224 format, which is a requirement for using the pre-trained VGG-16 CNN [33], and then processed into 4096 features for each input image. This number of features is the outcome of several intermediate processing steps taking place in the network's layers, each one returning a different number of outputs. As an example, the second layer has the following structure $(224, 224, 64)$, which means that is returns 64 images in the 224×224 format, as graphically shown in Fig.4. From this figure it can be seen how the CNN has already started picking up common patterns.

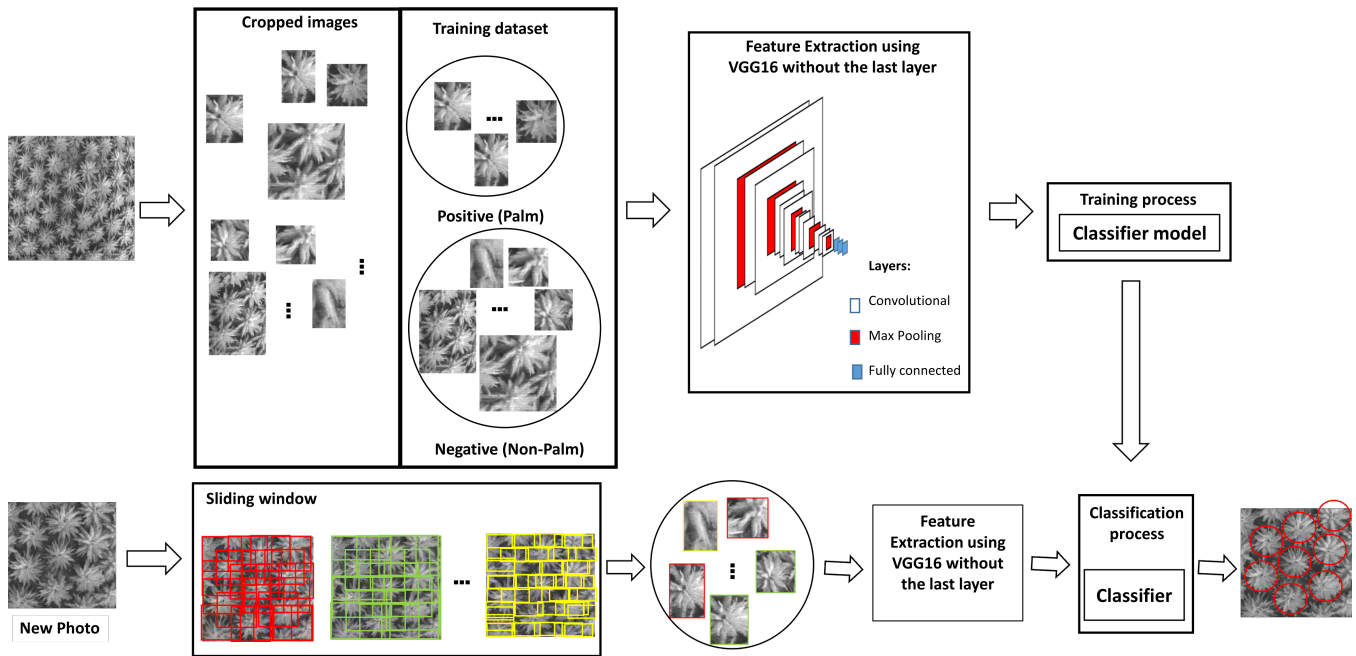3) **Training the classifier**: the totality of the features ob-

Fig. 3. A general overview of the proposed system, from preparing the training dataset for transfer learning to feature extractions and classification.

tained from the previous step, which are represented with numerical descriptors suitable for any classifier, form an accurate training set to fine tune the oil palm detection process.

In this study, the classifier is a SVM with linear kernel – a description of this method is given in section II-F. This was selected empirically amongst other tested classification methods based on: K-NN [43], [44], artificial neural networks [45], [46], decision trees [47], [48] and SVMs equipped with the non-linear kernel functions (refer to section II-F for details) which were compared against each other by using $10-$fold cross-validation and the evaluation metrics indicated in section II-G.

With reference to Fig. 3, these first three steps refer to the preliminary phase shown in the top row. Subsequently, the classification process can be executed with any new photo as input (bottom row).

4) **Oil palm unit detection**: for any new photo entering our proposed system, see bottom row of Fig. 3, the number of outputs (in terms of detection) depends on its content, i.e. the number of oil palm units in the image. These can be efficiently recognised with the previously trained classifier (i.e. step 3) only if fed with smaller versions of the original image (resembling the ones used for training). Indeed, no classifier wold be able to work optimally by considering the original image as a whole.

Thus, the sliding windows mechanism described in section II-C must be first employed to segment the input image and create a gallery containing copped portions from different central positions and with different sizes. This is followed by the feature extraction process taking place with the same methodology explained in step 2 (with the only difference that in step 2 features were used to train the classifier while in this step they represent the inputs to be classified).

Finally, the obtained features enter the classification stage, where the previously trained (i.e. at step 3) model is used to detect oil palms unit by making positive or negative decisions.

IV. RESULTS

The proposed system underwent the $10-$fold Cross-Validation (CV) process for which the corresponding confusion matrix values are displayed in Table III. These are used to evaluate the performances of the detection process in terms of the performance metrics defined in section II-G – a summary of the results is reported in Table IV.

TABLE III
CONFUSION MATRIX OF THE 10-FOLD CROSS-VALIDATION

		Predicted	
		Positive	Negative
Actual	**Positive**	351	5
	Negative	2	936

At a first glance, it is evident that the proposed classification process perform optimally with a more than satisfactory accuracy of 99.5%. It is worth noting that none of employed metrics has an inferior value, and with the precision metric displaying a performance peak of 99.8%. This points out the effectiveness of the selected methods for addressing these specific application domains in the field of precision agriculture.
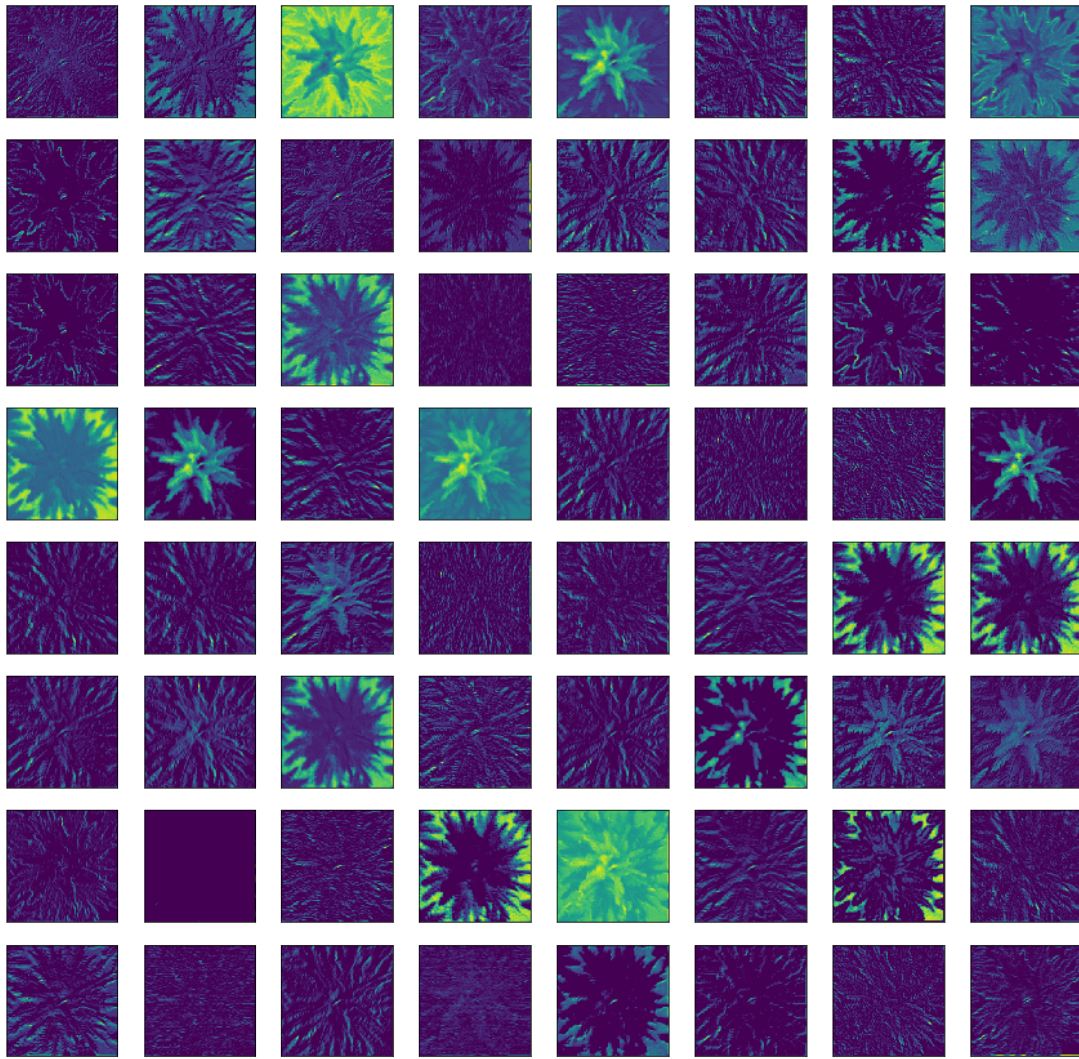
Fig. 4. Example of an output from the second layer of the employed pre-trained VGG-16 CNN.

To further highlight the suitability of our model for oil palm units detection, a second "validation" test was performed with a randomly chosen UAV image from those collected at a different altitude of 20m (which is displayed in Fig. 5). Fig. 5 shows a small section of that image used. As can be seen in the figure, in the image there are 6 palms to identify as individual plants. Those that were well classified are enclosed in a green rectangle, and a red one shows the palm that was not identified. It was expected that this was not identified because there is only part of the palm visible and the classifier is trained to recognise a palm only when it is complete.

Due to the different altitude, the maximum range for the side of the square sliding window, see section II-C, was adjusted to $277 - 469$. Since the classifier is trained with images taken at an altitude of 50m, one may expect a deterioration of the performances. However, results (shown in the second row of the Table IV) indicate a very good classification precision of 98.3% and slightly inferior values for accuracy,
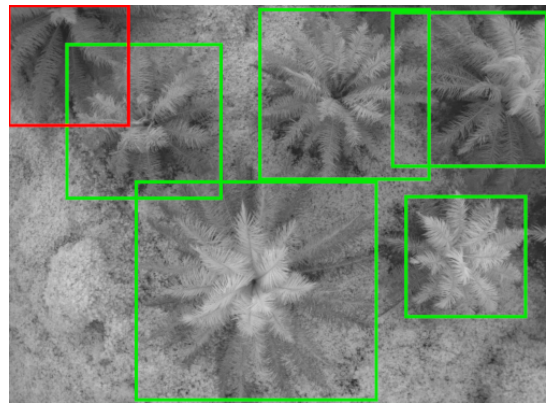

Fig. 5. Validation UAV photograph. Unlike the images in [27], this photographs was shot at an altitude of 20m.

F1 and recall. Hence, only a marginal deterioration of the classification performance is recorded, but the outcome is still very satisfactory.

TABLE IV
RESULTS OF THE PROPOSED MODEL WITH CROSS-VALIDATION 10-FOLD

Database	Accuracy	F1	Precision	Recall
Training data (CV)	99.5%	99.6%	99.8%	99.5%
Validation	97.5%	97.1%	98.3%	97.7%

It is important to note that this palm prediction problem, when doing it with the window technique, is going to have many sub-images that have the same palm or pieces of the same palm. This can give a wrong idea of the prediction percentage when we analyse the sub-images, increasing the number of miss-classified palms, when they were possibly already classified correctly by some window. This is why in future works it is recommended to do the validation on the number of palms detected per image.

V. CONCLUSION AND FUTURE WORK

In conclusion, the proposed system is highly effective, which leads to the following considerations:

- the use of TL is proven to be valid also in precision agriculture to pursue plants recognition;
- the jointly used of CNNs and SVM seems to be promising considering that very good results are ac hived by using a modification of a well-known VGG-16 (with the last layer removed) and a SVM with a simple linear kernel, without having to perform any modification on their algorithmic structures;
- in addition to being very accurate, the proposed approach is very flexible as it can be used straightaway for identifying similar plants, or a different variety if a new dataset is created for the training and validation;
- the proposed classification method has a high potential impact since this technology can be used in large plantations to identify e.g. diseased units, or units needing assisted pollination.

Despite the satisfactory results, we have identified some strategies to further improve the proposed system which will be investigated in the future. Some interesting points to be considered in our follow up research are:

- considering a higher number of modern classifiers and feature extraction methods (and their possible combination) might lead to better results or more flexible systems;
- to increase the number images in the dataset and prevent possible overfitting. Increasing the data will also help us, in addition to cross-validation, to be able to leave a validation set and use it to assess the performance of our prediction model;
- a smarter alternative to the sliding window technique which, unlike what happens in the current system, has a null probability of selecting the same palm twice will

result into a better classification and a faster system (i.e. classification repeats do not occur). A simple approach to solve this could consist in the modification of the window pitch when a palm is detected, so that the next window is positioned next the previous one without causing overlap.

REFERENCES

[1] T. T. Sue, *Oil palm uses*. Malaysian palm oil board (MPOB), 2017. [Online]. Available: www.mpob.gov.myi

[2] J. Berni, P. Zarco-Tejada, L. Suárez, V. González-Dugo, and E. Fereres, "Remote sensing of vegetation from uav platforms using lightweight multispectral and thermal imaging sensors," *Int. Arch. Photogramm. Remote Sens. Spatial Inform. Sci*, vol. 38, no. 6, p. 6, 2009.

[3] G. Prince, J. P. Clarkson, N. M. Rajpoot *et al.*, "Automatic detection of diseased tomato plants using thermal and stereo visible light images," *PloS one*, vol. 10, no. 4, p. e0123262, 2015.

[4] J. Meléndez, J. M. Aranda, A. J. de Castro, and F. López, "Measurement of forest fire parameters with multi-spectral imaging in the medium infrared," *Quantitative InfraRed Thermography Journal*, vol. 3, no. 2, pp. 183–199, 2006.

[5] C. Strecha, A. Fletcher, A. Lechner, P. Erskine, and P. Fua, "Developing species specific vegetation maps using multi-spectral hyperspatial imagery from unmanned aerial vehicles," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. I-3, pp. 311–316, 2012.

[6] J. Torres-Sánchez, F. López-Granados, A. I. De Castro, and J. M. Peña-Barragán, "Configuration and specifications of an unmanned aerial vehicle (uav) for early site specific weed management," *PloS one*, vol. 8, no. 3, p. e58210, 2013.

[7] K. Golhani, S. K. Balasundram, G. Vadamalai, and B. Pradhan, "A review of neural networks in plant disease detection using hyperspectral data," *Information Processing in Agriculture*, vol. 5, no. 3, pp. 354 – 371, 2018.

[8] P. Ahmadi, F. M. Muharam, K. Ahmad, S. Mansor, and I. Abu Seman, "Early detection of ganoderma basal stem rot of oil palms using artificial neural network spectral analysis," *Plant Disease*, vol. 101, no. 6, pp. 1009–1016, 2017.

[9] B. H. Trisasongko and D. Paull, "A review of remote sensing applications in tropical forestry with a particular emphasis in the plantation sector," *Geocarto International*, vol. 0, no. 0, pp. 1–23, 2018.

[10] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 152, pp. 166 – 177, 2019.

[11] K. L. Chong, K. D. Kanniah, C. Pohl, and K. P. Tan, "A review of remote sensing applications for oil palm studies," *Geo-spatial Information Science*, vol. 20, no. 2, pp. 184–200, 2017.

[12] K. F. A. Darras, M. D. Corre, G. Formaglio, A. Tjoa, A. Potapov, F. Brambach, K. T. Sibhatu, I. Grass, A. A. Rubiano, D. Buchori, J. Drescher, R. Fardiansah, D. Hölscher, B. Irawan, T. Kneib, V. Krashevska, A. Krause, H. Kreft, K. Li, M. Maraun, A. Polle, A. R. Ryadin, K. Rembold, C. Stiegler, S. Scheu, S. Tarigan, A. Valdés-Uribe, S. Yadi, T. Tscharntke, and E. Veldkamp, "Reducing fertilizer and avoiding herbicides in oil palm plantations—ecological and economic valuations," *Frontiers in Forests and Global Change*, vol. 2, p. 65, 2019.

[13] K. Thenmozhi and U. S. Reddy, "Crop pest classification based on deep convolutional neural network and transfer learning," *Computers and Electronics in Agriculture*, vol. 164, p. 104906, 2019.

[14] B. Neupane, T. Horanont, and N. D. Hung, "Deep learning based banana plant detection and counting using high-resolution red-green-blue (RGB) images collected from unmanned aerial vehicle (UAV)," *PLOS ONE*, vol. 14, no. 10, p. e0223906, oct 2019.

[15] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, Oct 2010.

[16] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI Global, 2010, pp. 242–264.

[17] Z. Qin, F. Yu, and C. X. Liu, Chenchen, "How convolutional neural networks see the world — a survey of convolutional neural network visualization methods," *Mathematical Foundations of Computing*, vol. 1, no. A0000-0001_2018_2_149, p. 149, 2018.

[18] H. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[19] W. Li, R. Dong, H. Fu, and L. Yu, "Large-scale oil palm tree detection from high-resolution satellite images using two-stage convolutional neural networks," *Remote Sensing*, vol. 11, no. 1, 2018.

[20] W. Li, H. Fu, L. Yu, and A. Cracknell, "Deep learning based oil palm tree detection and counting for high-resolution remote sensing images," *Remote Sensing*, vol. 9, no. 1, 2017.

[21] N. A. Mubin, E. Nadarajoo, H. Z. M. Shafri, and A. Hamedianfar, "Young and mature oil palm tree detection and counting using convolutional neural network deep learning method," *International Journal of Remote Sensing*, vol. 40, no. 19, pp. 7500–7515, 2019.

[22] A. Peña, I. Bonet, D. Manzur, M. Góngora, and F. Caraffini, "Validation of convolutional layers in deep learning models to identify patterns in multispectral images," in *2019 14th Iberian Conference on Information Systems and Technologies (CISTI)*, June 2019, pp. 1–6.

[23] Y. Wang, X. Zhu, and B. Wu, "Automatic detection of individual oil palm trees from uav images using hog features and an svm classifier," *International Journal of Remote Sensing*, vol. 40, no. 19, pp. 7356–7370, 2019.

[24] I. Bonet, F. Caraffini, A. Pena, A. Puerta Echandía, and M. Gongora, "Nir dataset for palm unit identification," Jan 2020. [Online]. Available: www.doi.org/10.21253/DMU.11743098

[25] "Unipalma," https://www.unipalma.com/, 2020-01-18.

[26] I. Bonet, F. Caraffini, A. Pena, A. Puerta Echandía, and M. Gongora, "Oil palm detection via deep transfer learning (collection)," Jan 2020. [Online]. Available: www.doi.org/10.21253/DMU.c.4832466

[27] ——, "NIR photographs of oil palm units," Jan 2020. [Online]. Available: www.doi.org/10.21253/DMU.11739303

[28] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, May 2004. [Online]. Available: https://doi.org/10.1023/B:VISI.0000013087.49260.fb

[29] J. Lee, I. Bang, and S. Yang, "Object detection with sliding window in images including multiple similar objects," in *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, Oct 2017, pp. 803–806.

[30] M. H. M. Noor, Z. Salcic, and K. I.-K. Wang, "Adaptive sliding window segmentation for physical activity recognition using a single tri-axial accelerometer," *Pervasive and Mobile Computing*, vol. 38, pp. 41 – 59, 2017.

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

[32] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, Aug 2018.

[33] F. Chollet *et al.*, "Keras," https://keras.io, 2015.

[34] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[35] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, Oct 2010.

[36] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big data*, vol. 3, no. 1, p. 9, 2016.

[37] D. Zhang, *Support Vector Machine*. Cham: Springer International Publishing, 2019, pp. 179–205.

[38] V. N. Vapnik, *Statistical Learning Theory*. Wiley-Interscience, 1998.

[39] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, ser. COLT '92. New York, NY, USA: Association for Computing Machinery, 1992, p. 144–152.

[40] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference and prediction*, 2nd ed. Springer, 2009. [Online]. Available: http://www-stat.stanford.edu/ tibs/ElemStatLearn/

[41] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[42] L. Wu, Z. Liu, T. Bera, H. Ding, D. A. Langley, A. Jenkins-Barnes, C. Furlanello, V. Maggio, W. Tong, and J. Xu, "A deep learning model to recognize food contaminating beetle species based on elytra fragments," *Computers and Electronics in Agriculture*, vol. 166, p. 105002, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168169918312821

[43] S. Zhang, X. Li, M. Zong, X. Zhu, and D. Cheng, "Learning k for knn classification," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 3, Jan. 2017.

[44] Y. Xu, Q. Zhu, Z. Fan, M. Qiu, Y. Chen, and H. Liu, "Coarse to fine k nearest neighbor classifier," *Pattern Recognition Letters*, vol. 34, no. 9, pp. 980 – 986, 2013.

[45] M. H. Rafiei and H. Adeli, " A New Neural Dynamic Classification Algorithm," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 12, pp. 3074–3083, Dec 2017.

[46] Y. CHITRE, A. P. DHAWAN, and M. MOSKOWITZ, "Artificial neural network based classification of mammographic microcalcifications using image structure features," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 07, no. 06, pp. 1377–1401, 1993.

[47] P. H. Swain and H. Hauska, "The decision tree classifier: Design and potential," *IEEE Transactions on Geoscience Electronics*, vol. 15, no. 3, pp. 142–147, July 1977.

[48] J. Dou, A. P. Yunus, D. T. Bui, A. Merghadi, M. Sahana, Z. Zhu, C.-W. Chen, K. Khosravi, Y. Yang, and B. T. Pham, "Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the izu-oshima volcanic island, japan," *Science of The Total Environment*, vol. 662, pp. 332 – 346, 2019.

[49] A. Pena, M. Gongora, I. Bonet, and F. Caraffini, "Intelligent system to improve the sustainability of oil palm crops through the construction of forecasting maps," https://doi.org/10.21253/DMU.11638095, Jan 2020.